Research Data Curation Program: Building Buzz at UC San Diego

Reid Otsuji
19th Conference of ASLI, 2016
New Orleans, LA
What is all the buzz about?

Part 1
Who we are . . .

What are we doing?

Why are we doing it?

Who are we doing it for?

Part 2
The Library Research Data Digital Collections

Buzzing Ahead . . .

Image: http://library.ucsd.edu/dc/collection/bb529166j
Research Data Curation Program: The Brief History

• 2 year pilot program beginning in 2011.

• Learn how researchers, information technologists, and librarians work together with data.

• In 2013, the library reorganized and the Research Data Curation Program was a new addition to the program structure.

Research Data Curation Services

- Making research data openly discoverable and accessible for the long term.

- Data Management Plans
  Resources and advice for constructing data management plans for grant proposals

- Permanent Data Identifiers
  Services for obtaining and managing long-term identifiers for digital content

- Research Data Curation
  Services for managing data throughout their life cycle

The Library UC San Diego
Who we are...now

Old team:

UCSD Libraries Team
- Aryds Kozbial, Director, Research Data Curation Program
- David Minor, Head of Curation and Preservation Services, San Diego Supercomputer Center
- Mary Linn Bergstrom, Curation Analyst and Faculty Liaison
- Sue McGuinness, Faculty Liaison
- Brad Westbrook, Metadata Consultant
- Declan Fleming, Technology Consultant

New team:

Research Data Curation Program Team

Our team is here to provide you with a variety of services and support.

Contact the Research Data Curation Program with any questions or comments. Feel free to email any of us individually if we can be of service.

David Minor is the Program Director for Research Data Curation. He oversees general direction for the group, and leads strategic partnerships on campus and within the UC. He’s also Director of the Chronopolis Program, a long-term digital preservation network. (ORCID, LinkedIn, ResearchGate)

Ho Jung Yoo is a Technical Analyst for Research Data Curation. She works directly with researchers seeking to manage and share their data. She brings to the program years of research experience in population ecology. (ORCID, ResearchGate)

Ryan Johnson is a Metadata Specialist for Research Data Curation. He helps researchers describe their data and makes sure it's discoverable in the digital repository. He has experience in managing metadata for digital humanities and social science data.

Julienne Schneider is a Research Data Strategist for Research Data Curation. She works with the processes involved in the data lifecycle, including the areas of discoverability, ontologies, data sharing and the federal requirements for data sharing and data management plans. Her experience includes development and use of ontologies in discoverability, development of tools for text mining, and data analysis/ transformation.

Mary Linn Bergstrom is a Liaison Librarian. She manages the EZID service, funded by the Library, whereby campus researchers can set up accounts and assign digital object identifiers (DOIs) to their research datasets. She also participates in developing and teaching workshops on the concepts of research data management and curation. (ORCID)

Reid Otting is a Data Curation Specialist and Faculty Liaison Librarian for Research Data Curation. He helps campus researchers with creating data management plans using the DMPTool and designs and teaches curriculum for data curation-related training and workshops. (ORCID)

Sibyl Schuyler is the Digital Preservation Analyst for Research Data Curation. She is Program Manager for Chronopolis and helps define long-term digital preservation solutions for the UC San Diego campus, also continuing our commitment to the Digital Preservation Network (DPN) and the National Digital Stewardship Alliance (NDSA). (LinkedIn)

Tim Dennis is the Library Data Librarian. He provides consultations on research involving data, including finding and recommending data sources and advising on technical data issues, such as file format conversion, web scraping, and statistical software use. He teaches workshops on various data and programming topics, including software use, data manipulation, management, and literacy. He also gives in-class presentations to introduce students to data sources and statistical software. (LinkedIn, ORCID)

http://libraries.ucsd.edu/services/data-curation/contact-us.html
What are we doing?

Building strategic partnerships.
What are we doing?

Promotion and Outreach.

- Hosting luncheon forums.
- Website redesign.
- Giving away promotional swag.
- Tabling at library sponsored events, new faculty, graduate, and post-doctoral orientations.
- Posting flyers on bulletin boards around campus.
- Networking at local research related events.
- Working with our liaisons librarians.
Developing training.

- Build data management training plan.

- Library Learning Services Program – developed by trained instructional designers.

- Instructional designers developed a comprehensive Learning Support Plan (LSP) for us to use as a resource for training development.
Improved Training, From this . . .

Initial training plan started with one-shot workshops for librarians and faculty researchers.

<table>
<thead>
<tr>
<th>Before Learning Support Plan (LSP)</th>
<th>Audiences</th>
<th>Training Support Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Faculty Researchers</td>
<td>Learning Resources</td>
</tr>
<tr>
<td>Data Management</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

Learn more at: [DMPTool](#)
Instructional designers created a Learning Support Plan matrix that outlined where we should focus our resources.

<table>
<thead>
<tr>
<th>Tasks and Processes</th>
<th>Audiences</th>
<th>Learning Support Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Librarians</td>
<td>Reference Resources</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Job Aids (Steps)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Policies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online Demos</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Workshops</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consultations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>People Resources</td>
</tr>
</tbody>
</table>

- **Use RDCP Services and Resources to Apply Best Practices of Data Management**

**Use RDCP Services and Resources to Apply Best Practices of Data Management**

<table>
<thead>
<tr>
<th>Reference Resources</th>
<th>Learning Resources</th>
<th>People Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAQ</td>
<td>Job Aids (Steps)</td>
<td>Policies</td>
</tr>
</tbody>
</table>

**Audiences**
- Librarians
- Researchers
- Faculty
- Students
- Graduate
- Undergraduate

**Concepts**

1. Understand the data management lifecycle
2. Create a Data Management Plan
3. Understand preservation and the value of easy access sharing
4. Establish workflows
5. Understand funder mandates
6. Understand storage options

- **1. Navigate website**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **2. Establish DMP tool account**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **3. Create EZID**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **4. Create naming convention**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **5. Create file structure**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **6. Create metadata**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **7. Create DOIs/ORCID**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **8. Deposit data**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **9. Use storage depository (DAMS)**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •

- **10. Use long term storage (Chronopolis)**
  - Librarians: •
  - Researchers: •
  - Faculty: •
  - Students: •
  - Graduate: •
  - Undergraduate: •
To This . . . Introducing Data Analysis Tools!

Exploring the need to teaching popular tools used in data science for analyzing and cleaning data.

https://software-carpentry.org/

https://software-carpentry.org/

http://www.datacarpentry.org/

GitLab

GitHub

Python

R

RStudio

The Library UC San Diego
Data Management 101 - School of Global Policy and Strategy
Feb 17- Mar 2, 2016
12:30 pm -1:50 pm

Instructors: Juliane Schneider, Reid Otsuji, Tim Dennis, Hyeonsu Kang
Helpers: Reid Otsuji, Tim Dennis

General Information
This is the website for the School of Global Policy and Strategy short course in data management and SQL. This course will introduce you best practices in data management. In order to earn a certificate of proficiency, you must attend all of the class meetings for the course, do the short coding assignments, and pass the short, in-class quiz at the end of the course. Courses begin on Monday, January 4th. This is the first day of instruction, winter quarter.

Who: The course is aimed at GPS graduate students. You don't need to have any previous knowledge of the tools that will be presented at the workshop.

Where: Room 3202, 9500 Gilman Drive, #0519, La Jolla, CA. Get directions with OpenStreetMap or Google Maps.

Requirements: Participants must bring a laptop with a few specific software packages installed (listed below).

Contact: Please mail timdennis@ucsd.edu for more information.

Assessment: A certificate of proficiency will be given to students who:
- Attends all classes
- Passes all in-course quizzes
- Satisfactorily completes the weekly assignments

Need help?: Email timdennis@ucsd.edu or schedule an appointment

Why are we doing it?

- Management of data throughout its lifecycle ...  

... to ensure it is reliably retrievable for future research purposes or reuse.
Who are we doing it for?
Data is the intellectual capital of the modern research university. It belongs somewhere it can be kept safely and discoverable.
Browsing the Digital Collections

http://library.ucsd.edu/dc/rdcp/collections

Browse by Collection: Library Digital Collections

View current collections below. Click an image or link for a full description.

Search Digital Collections

Baja California
Images of Baja California digitized from original photographs held by UC San Diego Library. Includes landscapes, missions, the built environment, and ranch life. 1938-1997

Bear Facts
Bear Facts was the newsletter published by the Oceanids, a service and social organization promoting friendship and services to the UC San Diego community. 1962-2007

California Explores the Ocean
A collection of diverse resources about the ocean and ocean exploration.

Camp Matthews Photographs and Plans
Photographs and plans of Camp Matthews buildings after Camp Matthews closed operations. 1964-1967

Charles Atwood Kofoid Papers
Photographs taken and collected by the University of California, Berkeley zoologist Charles Atwood Kofoid and his wife, Carrie Winter Kofoid. Subjects include the Kofoid's; their families, friends and colleagues; travel; the 1904-1905 Albatross Expedition; and many freshwater and marine biological research stations. circa 1890-1936

Charles H. Graves Photographs and Films
Film and color slides created for and used in a photographic documentary of early UCSD events. 1963-1967

Chinese Cultural Revolution Posters
Propaganda posters created during the Cultural Revolution by various Chinese agencies. 1966-1989

http://library.ucsd.edu/dc/rdcp/collections

Browse by Collection: UCSD Research Data Collections

View current collections below. Click an image or link for a full description.

Search Digital Collections

Results 1 - 12 of 12

An annotated checklist of the bees (Hymenoptera: Anthophila) of San Diego County, California
This database represents an ongoing survey to document all the bee species that occur within the bounds of San Diego County. As this project is still ongoing, the database will be updated periodically with additions and/or corrections. July 2015,

Bee Research Methods: Video Demonstrations
This collection shows bee behaviors and methods used to study bee behaviors. 2005-2014

California Tobacco Survey (CTS)
The objective of these surveys is to collect representative statewide data on cigarette smoking behavior, attitudes towards smoking, media exposure to smoking, and use of tobacco products other than cigarettes, from populations living in California. 1990-

Data from: Allometry through the Computational Microscope: cAMP Activation of a Canonical Signalling Domain
This collection contains the molecular dynamics trajectories used in the paper "Allometry through the computational microscope: cAMP activation of a canonical signalling domain". 2014-08-24,

Data from: Carbonic Anhydrases, EPF2 and a Novel Protease Mediate CO2 Control of Stomatal Development
RNA-Seq experiments analyzing the Arabidopsis response to low and elevated carbon dioxide growing conditions were conducted to capture early signaling events as seedlings were adapting to stress. 2012-2013

Data from: Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning
The kinetic rate constants of four biochemically relevant ligand-receptor systems were estimated and described in a publication. This collection includes input files, output files, trajectories, and statistics associated with the study. 2013-2015

http://library.ucsd.edu/dc/rdcp/collections
<table>
<thead>
<tr>
<th>Collection</th>
<th>Items Count</th>
<th>Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Khirbat en-Nahas Project&quot;@ar-Latn</td>
<td>1366</td>
<td>207,540</td>
</tr>
<tr>
<td>An annotated checklist of the bees (Hymenoptera: Anthophila) of San Diego County, California</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bee Research Methods: Video Demonstrations</td>
<td>5</td>
<td>1,068</td>
</tr>
<tr>
<td>California Tobacco Survey (CTS)</td>
<td>9</td>
<td>4,137</td>
</tr>
<tr>
<td>Data from: Allostery through the Computational Microscope: cAMP Activation of a Canonical Signalling Domain</td>
<td>1</td>
<td>19,686</td>
</tr>
<tr>
<td>Data from: Carbonic Anhydrases, EPF2 and a Novel Protease Mediate CO2 Control of Stomatal Development</td>
<td>1</td>
<td>31,005</td>
</tr>
<tr>
<td>Data from: Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning</td>
<td>1</td>
<td>268,703</td>
</tr>
<tr>
<td>Heavy Metals in the Ocean Insect, Halobates</td>
<td>2</td>
<td>92</td>
</tr>
<tr>
<td>Keith Rayner Eye Movements in Reading Data Collection</td>
<td>15</td>
<td>4,351</td>
</tr>
<tr>
<td>OpenTopography</td>
<td>103</td>
<td>4,810,330</td>
</tr>
<tr>
<td>SWell Ex-96 Experiment Acoustic Data</td>
<td>4</td>
<td>3,258</td>
</tr>
<tr>
<td>Santa Fe Light Cone Simulation research project files</td>
<td>39</td>
<td>751,579</td>
</tr>
<tr>
<td>Scripps Institution of Oceanography, Geological Collections</td>
<td>6761</td>
<td>4,798</td>
</tr>
<tr>
<td>TeachingBee</td>
<td>12</td>
<td>259</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td><strong>8,305</strong></td>
<td><strong>6,102,455</strong></td>
</tr>
</tbody>
</table>
The Bees Got Us Started

Teaching Bee

About this collection

Description
This collection of documents provides teaching exercises and information for instructors of students ranging from junior high school through college. The target audience is indicated for each collection item.

Creator
James Nieh

Creation Date
2005-2013

Extent
12 digital objects.

Formats

- video
- text

http://library.ucsd.edu/dc/collection/bb6725113c
The Bees Got Us Started

Multimedia capable.

Downloadable files.
Heavy Metals in the Ocean Insect, Halobates

About this collection

Description

Halobates is the only insect genus with representatives living their entire lives in the open ocean. It is a member of the true bug order Heteroptera and belongs to the family Gerridae which includes pond skaters commonly found in freshwater ponds, lakes, streams and rivers. Although the genus was first discovered in 1822, not much was known about its biology or special adaptations which enabled it to live in the open ocean where no other insects were able to survive. This was largely because few entomologists have any reasons to venture out to sea. Being attached to the Scripps Institution of Oceanography (SIO) with sea-going vessels has given me opportunities to study them at sea. What we now know about their distributions, biology, special adaptations and phylogeny can be found in reviews listed under 'Publication'.

This project was initiated when interests on heavy metal pollution Worldwide were generated following discoveries on lead poisoning from additives in petrol in the 1970s. Air pollutants eventually rain down to earth. Since the ocean covers more than 70% of the earth's surface, many chemical oceanographers were interested in finding out the presence and concentrations of heavy metals in the sea. It is relatively straightforward to measure heavy metal concentrations in seawater but measuring those that occur at the surface film which is the entry point of pollutants to the ocean is quite another matter. Methods and precautions needed in order to prevent contamination of surface samples were difficult to surmount. Halobates are completely wingless. They do not fly, nor dive throughout their lives. They are completely restricted to the sea-air Interface. They were easy to collect and samples could be cleaned prior to measurements to give dependable results. They can therefore be useful indicators of heavy metal concentrations on the ocean film.

Principal Investigator

Cheng, Lanna

Co Principal Investigator

Schulz-Baldes, Meinhard

Contributors

Alexander, George V.
Franco, Paul J.
Ott, John

Date Issued

2015

Date Collected

1961-1978

http://library.ucsd.edu/dc/collection/bb8056206n
Unique Dataset

Heavy Metals in the Ocean Insect, Halobates

- Scanned 63 computer printouts from the early 70’s.
- OCR completed and verified.
- Manually verified and tested Excel data entries

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>sequence</td>
<td>B</td>
<td>specimen</td>
<td>C</td>
<td>species</td>
</tr>
<tr>
<td>2</td>
<td>34 033</td>
<td>SE</td>
<td>20N</td>
<td>119W</td>
<td>7.721</td>
</tr>
<tr>
<td>3</td>
<td>35 034</td>
<td>SE</td>
<td>-19S</td>
<td>-105W</td>
<td>4.439</td>
</tr>
<tr>
<td>4</td>
<td>36 035</td>
<td>SE</td>
<td>20N</td>
<td>-119W</td>
<td>7.044</td>
</tr>
<tr>
<td>5</td>
<td>37 036</td>
<td>SE</td>
<td>-13S</td>
<td>-126W</td>
<td>3.049</td>
</tr>
<tr>
<td>6</td>
<td>38 037</td>
<td>SE</td>
<td>-16S</td>
<td>-126W</td>
<td>53.431</td>
</tr>
<tr>
<td>7</td>
<td>39 038</td>
<td>SE</td>
<td>-15S</td>
<td>-126W</td>
<td>13.526</td>
</tr>
<tr>
<td>8</td>
<td>40 039</td>
<td>SE</td>
<td>-17S</td>
<td>-126W</td>
<td>9.722</td>
</tr>
<tr>
<td>9</td>
<td>41 040</td>
<td>SE</td>
<td>-20S</td>
<td>-125W</td>
<td>6.919</td>
</tr>
<tr>
<td>10</td>
<td>42 041</td>
<td>SE</td>
<td>-20S</td>
<td>-126W</td>
<td>9.575</td>
</tr>
<tr>
<td>11</td>
<td>43 042</td>
<td>SE</td>
<td>-20S</td>
<td>-125W</td>
<td>13.803</td>
</tr>
<tr>
<td>12</td>
<td>44 043</td>
<td>SE</td>
<td>-20S</td>
<td>-126W</td>
<td>3.930</td>
</tr>
<tr>
<td>13</td>
<td>45 044</td>
<td>SE</td>
<td>20N</td>
<td>-126W</td>
<td>13.384</td>
</tr>
<tr>
<td>14</td>
<td>46 045</td>
<td>SE</td>
<td>-15S</td>
<td>-119W</td>
<td>10.561</td>
</tr>
<tr>
<td>16</td>
<td>48 047</td>
<td>SE</td>
<td>-18S</td>
<td>-119W</td>
<td>8.268</td>
</tr>
<tr>
<td>17</td>
<td>49 048</td>
<td>SE</td>
<td>-20S</td>
<td>-122W</td>
<td>8.268</td>
</tr>
<tr>
<td>18</td>
<td>123 120</td>
<td>SE</td>
<td>12N</td>
<td>-145W</td>
<td>1.510</td>
</tr>
<tr>
<td>19</td>
<td>159 157</td>
<td>SE</td>
<td>21N</td>
<td>-158W</td>
<td>0.687</td>
</tr>
<tr>
<td>20</td>
<td>160 158</td>
<td>SE</td>
<td>21N</td>
<td>-158W</td>
<td>2.275</td>
</tr>
<tr>
<td>21</td>
<td>166 164</td>
<td>SE</td>
<td>21N</td>
<td>-151W</td>
<td>5.756</td>
</tr>
<tr>
<td>22</td>
<td>172 170</td>
<td>SE</td>
<td>17N</td>
<td>-151W</td>
<td>1.627</td>
</tr>
<tr>
<td>23</td>
<td>173 171</td>
<td>SE</td>
<td>17N</td>
<td>-151W</td>
<td>2.396</td>
</tr>
<tr>
<td>24</td>
<td>176 174</td>
<td>SE</td>
<td>19N</td>
<td>-156W</td>
<td>1.235</td>
</tr>
<tr>
<td>25</td>
<td>178 176</td>
<td>SE</td>
<td>19N</td>
<td>-156W</td>
<td>0.654</td>
</tr>
<tr>
<td>26</td>
<td>179 177</td>
<td>SE</td>
<td>19N</td>
<td>-156W</td>
<td>1.298</td>
</tr>
<tr>
<td>27</td>
<td>180 178</td>
<td>SE</td>
<td>21N</td>
<td>-158W</td>
<td>9.504</td>
</tr>
<tr>
<td>28</td>
<td>181 179</td>
<td>SE</td>
<td>21N</td>
<td>-158W</td>
<td>0.600</td>
</tr>
<tr>
<td>29</td>
<td>182 180</td>
<td>SE</td>
<td>19N</td>
<td>-156W</td>
<td>1.051</td>
</tr>
<tr>
<td>30</td>
<td>183 181</td>
<td>SE</td>
<td>21N</td>
<td>-159W</td>
<td>8.277</td>
</tr>
</tbody>
</table>
Handwritten collection event record was added as a part of the dataset.
Buzzing Ahead

- Improve existing services and Increase the number of research data curated by the library.

- Further streamlining the data collection workflow.

- Continue pursuing partnerships with academic departments and researchers through outreach with our liaison librarians.

- Implement new training ideas and workshops to encourage more use of our tools and services.
Thank you!

Questions?

Reid Otsuji
Data Curation Specialist and Faculty Liaison
UC San Diego Library Research Data Curation Program
rotsuji@ucsd.edu

UC San Diego Library Digital Collections
http://library.ucsd.edu/dc

The Library Research Data Curation Program, UC San Diego
http://lib.ucsd.edu/rdcp